Scalable approach to quantum simulation via projection-based embedding

Alexis Ralli[1,*], Michael Williams de la Bastida[1,†] and Peter V. Coveney[1,2,3,‡]

[1]*Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom*
[2]*UCL Centre for Advanced Research Computing, Gower Street, London WC1E 6BT, United Kingdom*
[3]*Informatics Institute, University of Amsterdam, Amsterdam 1098 XH, Netherlands*

Owing to the computational complexity of electronic structure algorithms running on classical digital computers, the range of molecular systems amenable to simulation remains tightly circumscribed even after many decades of work. Many believe quantum computers will transcend such limitations although in the current era the size and noise of these devices militates against significant progress. Here we describe a chemically intuitive approach that permits a subdomain of a molecule's electronic structure to be calculated accurately on a quantum device, while the rest of the molecule is described at a lower level of accuracy using density functional theory running on a classical computer. We demonstrate that this approach produces improved results for molecules that cannot be simulated fully on current quantum computers but which can be resolved classically at a cheaper level of approximation. The algorithm is tunable, so that the size of the quantum simulation can be adjusted to run on available quantum resources. Therefore, as quantum devices become larger, this method will enable increasingly large subdomains to be studied accurately.

I. INTRODUCTION

Quantum computing is anticipated to enable accurate simulation of chemical systems beyond the capabilities of classical methods. Whether this aim will be achieved with so-called noisy intermediate-scale quantum (NISQ) processors is still to be seen [1–4]. While devices are improving rapidly, NISQ applications also require algorithmic tools to mitigate noise and reduce required qubit counts.

Embedding procedures work by first partitioning a system and then applying differing levels of theory to each region. An accurate but computationally expensive method is applied to a small *active region* [5,6]. The surrounding *environment* is handled with a more efficient but approximate method. This allows some of the physically relevant detail to be captured while avoiding the computational cost of accurately simulating the entire system. However, even for fairly small active regions, exact classical simulation using the full configuration interaction (FCI) method quickly becomes unfeasible due problem scaling exponentially with system size [7]. For such problems, the number of Slater determinants scales as $\binom{M}{N}$, for N electrons and M orbitals [4].

The current "gold standard" in conventional quantum chemistry is coupled cluster (CC) theory, which offers a good accuracy-to-cost ratio and reduces this factorial complexity [8,9]. The CC singles and doubles (CCSD) method scales as $O(M^6)$ [10]. The CCSD(T), which treats the triple excitations perturbatively, scales as $O(M^7)$ in time [4]. This still imposes practical limitations on system size while imperfectly approximating the effects of correlation [11]. Therefore, classical embedding methods still inevitably inherit the shortcomings of such methods, even within a smaller active region. In short, accurately simulating quantum effects at large scale remains elusive.

Quantum computers can efficiently represent the state of general quantum systems and provide a practical way to perform quantum chemistry simulations in polynomial time [12]; however, this approach will only be possible in the fault-tolerant regime, as it requires the quantum phase estimation (QPE) algorithm [13] which cannot be implemented on current NISQ quantum computers [14,15]. There are, however, still open questions regarding the advantage of using quantum computers for chemistry simulations [16]. Quantum algorithms designed for NISQ devices, such as the variational quantum eigensolver (VQE) [17], allow quantum systems to be studied using present-day hardware; however, this is limited by the current quality and quantity of qubits. To date, the largest chemical VQE simulations are a 12-qubit study of a H_{12} chain [18] and a 12-qubit simulation of Li_2O [19].

By embedding a wave-function simulation calculated on a quantum computer into a larger classical simulation, we can mitigate some of the shortfalls of classical hardware in describing quantum systems, while requiring fewer qubits and shorter quantum circuits than full-system quantum simulation. This will allow systems normally too large to study at the wave-function level to be modeled via a multiscale approach. In this way, embedding can serve as an algorithmic tool to

*alexis.ralli.18@ucl.ac.uk
†michael.williams.20@ucl.ac.uk
‡p.v.coveney@ucl.ac.uk

mitigate the shortcomings of quantum and classical processors, thereby providing novel results. Additionally, as embedding methods may utilize fault-tolerant quantum simulation methods, they will continue to facilitate the study of systems larger than would otherwise be possible. Hybrid embedding methods published to date include wave function in density functional theory (DFT) [20,21], density matrix embedding theory (DMET) [22–24], and Dynamical Mean-Field Theory (DMFT) [25–27] approaches.

We reformulate the projection-based embedding method, which enables the application of quantum algorithms to molecules of arbitrary size while consistently improving on the results of full-system DFT. This method outputs a Hamiltonian which can be solved using any suitable NISQ or fault-tolerant quantum algorithm, thus augmenting the usefulness of quantum processors in general. We anticipate that by targeting quantum processors at regions with strong correlation, hybrid embedding will enable novel results. This concept has been utilized in other works, where a smaller embedded qubit Hamiltonian is produced and solved; however, the underlying theory and assumptions are different for each hybrid model [20,28,29]. The method presented in this work is distinct from these.

II. BACKGROUND

In this section, we cover the necessary background on the projection-based embedding technique and discuss the molecular orbital (MO) localization methods considered in this work.

A. Projection-based embedding

Quantum embedding schemes for electronic structure problems aim to reduce the computational cost of a problem by dividing a molecular problem into smaller (and so less-costly) subsystems [30]. In this paper, we employ the projection-based embedding (PBE) method, first proposed by Manby *et al.* [31], that offers a practical way to implement formally exact quantum embedding. At a high level, a molecular problem is split into an active and an environment region. The active system is then solved at a more accurate (and thus computationally more expensive) level of theory than the environment. The PBE method allows rigorous embedding of either a wave-function (WF) subsystem into a self-consistent field (SCF) environment (WF-in-SCF embedding) or an SCF subsystem description in an SCF environment (SCF-in-SCF embedding). Here the SCF method should be thought of as Hartree-Fock or Kohn-Sham DFT (KS-DFT). We summarize the important details of this model here, where the environment SCF calculation is performed using KS-DFT.

To begin, an initial KS-DFT calculation of the entire system is carried out using a low (cheap) level of theory. This yields a set of MOs $\{\psi_i(\vec{r})|i = 1, 2, \ldots, N\}$. Each MO is formed from a linear combination of K known atomic orbital (AO) basis functions $\{\phi_j(\vec{r})|j = 1, 2, \ldots, K\}$:

$$\psi_i(\vec{r}) = \sum_{j=1}^{K} \mathbf{C}_{ji}\phi_j(\vec{r}), \tag{1}$$

where $\mathbf{C}$ is a matrix of MO coefficients. In general, the AO basis functions $\phi_j(\vec{r})$ are not orthonormal. We can see this by the $(K \times K)$ overlap matrix:

$$\mathbf{S}_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle = \int d\vec{r}\, \phi_\mu(\vec{r})^* \phi_\nu(\vec{r}). \tag{2}$$

If $\mathbf{S}$ is the identity matrix then all the AO basis functions are orthonormal; however, in general this is not the case. However, linear combinations of these nonorthogonal AOs, given by the columns of $\mathbf{C}$, construct orthogonal MOs $\psi_i(\vec{r})$, i.e., $\mathbf{C}^\dagger \mathbf{S}\mathbf{C} = \mathbf{I}$.

In order to partition the molecular problem into active and environment parts, these canonical MOs ψ_i must first be localized and assigned to a subsystem. This can be done by different localization methods (described in further detail in Appendix A). In effect, we use a unitary transform $\mathbf{U}$ (defined by the localization procedure) to spatially localize each ψ_i as much as possible. The reasoning behind this is introduced next.

Lehtola and Jónsson noted that "total energy in both Hartree-Fock and KS-DFT is invariant under a unitary transformation of the occupied-occupied and virtual-virtual blocks" [32]. This is true because the Hartree-Fock and KS-DFT wave functions are approximated as a single Slater determinant. From linear algebra, it is well known that the determinant of a matrix product is given by the product of their determinants, i.e., for general matrices $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$. Using this property and the fact that the determinant of the identity matrix is $\det(\mathbf{I}) = 1$, given a unitary matrix $\mathbf{V}$, where $\mathbf{V}^\dagger \mathbf{V} = I$, we obtain the following:

$$
\begin{aligned}
1 = \det(\mathbf{I}) &= \det(\mathbf{V}^\dagger \mathbf{V}) = \det(\mathbf{V}^\dagger)\det(\mathbf{V}) \\
&= \det(\mathbf{V})^* \det(\mathbf{V}) = |\det(\mathbf{V})|^2.
\end{aligned} \tag{3}
$$

This implies that the determinant of any unitary matrix must have a value of $e^{i\theta}$, as $|e^{i\theta}| = 1$. Therefore, acting with a unitary will leave the Slater determinant unchanged up to a global phase and so observable quantities of the wave function will be unchanged. Thus the solution of an SCF problem can be described by a set of different (unitarily) rotated orbitals. Such a unitary rotation $\mathbf{U}$ can be used to spatially localize each MO ψ_i as much as possible. The form of $\mathbf{U}$ is defined by a particular localization procedure and there are many methods based on different localization criteria. We denote these orbitals as localized molecular orbitals (LMOs) or ψ_i^{LMO}. The matrix of orbital coefficients for these localized orbitals is given by the columns of $\mathbf{C}^{\text{LMO}}$ defined as

$$\mathbf{C}^{\text{LMO}} = \mathbf{CU}. \tag{4}$$

This construction ensures the orthonormality condition of each molecular orbital is still conserved, i.e., $(\mathbf{C}^{\text{LMO}})^\dagger \mathbf{S}\mathbf{C}^{\text{LMO}} = \mathbf{I}$. We see this via the following proof:

$$
\begin{aligned}
(\mathbf{C}^{\text{LMO}})^\dagger \mathbf{S}\mathbf{C}^{\text{LMO}} &= \mathbf{U}^\dagger \mathbf{C}^\dagger \mathbf{S}\mathbf{C}\mathbf{U} \\
&= \mathbf{U}^\dagger \mathbf{I}\mathbf{U} = \mathbf{I}\mathbf{U}^\dagger \mathbf{U} \\
&= \mathbf{I}.
\end{aligned} \tag{5}
$$

This construction allows for $\mathbf{U}$ to be determined from $\mathbf{C}$ and $\mathbf{C}^{\text{LMO}}$:

$$(\mathbf{C}^{\text{LMO}})^{\dagger} \mathbf{S} \mathbf{C}^{\text{LMO}} = \mathbf{U}^{\dagger} \mathbf{C}^{\dagger} \mathbf{S} \mathbf{C}^{\text{LMO}} = \mathbf{I}$$
$$\Rightarrow \mathbf{U} = \mathbf{C}^{\dagger} \mathbf{S} \mathbf{C}^{\text{LMO}} \tag{6}$$

by multiplying on the left with $\mathbf{U}$. The reason we include Eq. (6) is that sometimes quantum chemistry packages only return $\mathbf{C}^{\text{LMO}}$ without $\mathbf{U}$.

In summary, we get the following mapping from canonical to localized molecular orbitals:

$$\psi_i(\vec{r}) = \sum_{j=1}^{K} \mathbf{C}_{ji} \phi_j(\vec{r}) \mapsto \psi_i^{\text{LMO}} = \sum_{j=1}^{K} [\mathbf{C}\mathbf{U}]_{ji} \phi_j(\vec{r})$$
$$= \sum_{j=1}^{K} \mathbf{C}_{ji}^{\text{LMO}} \phi_j(\vec{r}). \tag{7}$$

Next we show how the charge density remains unchanged.

For a closed-shell molecule, described by a single determinant wave function, each MO ψ_i contains two electrons and thus the total charge density is [7]

$$\rho(\vec{r}) = 2 \sum_{i=1}^{N/2} \psi_i^*(\vec{r}) \psi_i(\vec{r})$$
$$= 2 \sum_{i=1}^{N/2} \left(\underbrace{\sum_{\nu=1}^{K} \mathbf{C}_{\nu i}^* \phi_\nu^*(\vec{r})}_{\psi_i^*(\vec{r})} \underbrace{\sum_{\mu=1}^{K} \mathbf{C}_{\mu i} \phi_\mu(\vec{r})}_{\psi_i(\vec{r})} \right)$$
$$= \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} \left[2 \sum_{i=1}^{N/2} \mathbf{C}_{\mu i} \mathbf{C}_{\nu i}^* \right] \phi_\nu^*(\vec{r}) \phi_\mu(\vec{r})$$
$$= \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} \gamma_{\mu\nu} \phi_\nu^*(\vec{r}) \phi_\mu(\vec{r}). \tag{8}$$

Here the square brackets define the density matrix $\gamma_{\mu\nu}$ (defined in the AO basis),

$$\gamma_{\mu\nu} = 2 \sum_{i=1}^{N/2} \mathbf{C}_{\mu i} \mathbf{C}_{\nu i}^{\dagger}, \tag{9}$$

that for a set of basis function $\{\phi_j(\vec{r}) | j = 1, 2, \ldots, K\}$ fully specifies the charge density $\rho(\vec{r})$ [7]. The sum runs over $N/2$, as these are the occupied MOs of a closed-shell calculation. The whole matrix can be obtained as $\gamma^{\text{total}} = 2\mathbf{C}_{\text{occ}}(\mathbf{C}_{\text{occ}})^{\dagger}$, where occ denotes only using the occupied columns of the $\mathbf{C}$ matrix [the first $N/2$ columns, indexed by i in Eq. (9)]. In the localized basis, the density matrix remains unchanged as

$$\gamma^{\text{total}} = 2\mathbf{C}_{\text{occ}}(\mathbf{C}_{\text{occ}})^{\dagger}$$
$$= 2\left[\mathbf{C}_{\text{occ}}^{\text{LMO}} \mathbf{U}^{\dagger}\right] \left[\mathbf{U}(\mathbf{C}_{\text{occ}}^{\text{LMO}})^{\dagger}\right] \tag{10}$$
$$= 2\mathbf{C}_{\text{occ}}^{\text{LMO}} (\mathbf{C}_{\text{occ}}^{\text{LMO}})^{\dagger}.$$

Given a set of localized molecular orbitals, we partition them into two subsystems denoted act (active) and env (environment). There are different methods to do so and we summarize our approach in Appendix A. Overall we generate a set of (occupied) LMO indices $\mathcal{K}$ and $\mathcal{L}$ for the active and environment subsystems, respectively. The resulting charge density for each subsystem can then be written as

$$\gamma_{\mu\nu}^{\text{act}} = 2 \sum_{k \in \mathcal{K}} \mathbf{C}_{\mu k}^{\text{LMO}} (\mathbf{C}_{\nu k}^{\text{LMO}})^{\dagger}, \tag{11a}$$

$$\gamma_{\mu\nu}^{\text{env}} = 2 \sum_{l \in \mathcal{L}} \mathbf{C}_{\mu l}^{\text{LMO}} (\mathbf{C}_{\nu l}^{\text{LMO}})^{\dagger}, \tag{11b}$$

for closed-shell calculations. The set $\mathcal{K} \cup \mathcal{L}$ contains all the occupied molecular orbital indices.

The total system electron density is written as a sum of subsystem densities:

$$\gamma^{\text{total}} = \gamma^{\text{act}} + \gamma^{\text{env}}$$
$$= 2\mathbf{C}_{\mathcal{K}}^{\text{LMO}} (\mathbf{C}_{\mathcal{K}}^{\text{LMO}})^{\dagger} + 2\mathbf{C}_{\mathcal{L}}^{\text{LMO}} (\mathbf{C}_{\mathcal{L}}^{\text{LMO}})^{\dagger} \tag{12}$$
$$= 2\mathbf{C}_{\text{occ}}^{\text{LMO}} (\mathbf{C}_{\text{occ}}^{\text{LMO}})^{\dagger}.$$

The number of electrons will also be split according to $n_e^{\text{total}} = n_e^{\text{act}} + n_e^{\text{env}} = \text{Tr}(\mathbf{S}\gamma^{\text{act}}) + \text{Tr}(\mathbf{S}\gamma^{\text{env}}) = \text{Tr}(\mathbf{S}\gamma^{\text{total}})$, where Tr denotes the trace operation.

The energy of the full system can be found from its components via [33]

$$E[\gamma^{\text{act}}, \gamma^{\text{env}}] = \underbrace{\text{Tr}(\gamma^{\text{act}} \mathbf{h}_{\text{core}}) + \mathbf{g}(\gamma^{\text{act}})}_{\text{energy of isolated act system}}$$
$$+ \underbrace{\text{Tr}(\gamma^{\text{env}} \mathbf{h}_{\text{core}}) + \mathbf{g}(\gamma^{\text{env}})}_{\text{energy of isolated env system}}$$
$$+ \underbrace{\mathbf{g}(\gamma^{\text{act}}, \gamma^{\text{env}})}_{\text{nonadditive two-electron energy}}. \tag{13}$$

Here $\mathbf{h}_{\text{core}}$ is the one-electron core Hamiltonian and $\mathbf{g}$ groups the two-electron terms—Coulomb and exchange for Hartree-Fock and exchange-correlation for DFT. The nonadditive two-electron energy is given by

$$\mathbf{g}(\gamma^{\text{act}}, \gamma^{\text{env}}) = \mathbf{g}(\gamma^{\text{act}} + \gamma^{\text{env}}) - \mathbf{g}(\gamma^{\text{act}}) - \mathbf{g}(\gamma^{\text{env}}), \tag{14}$$

and accounts for the interaction between subsystems [33].

Next we want to solve the active system using a higher (more accurate) level of theory. The effect of the interaction between the active and environment subsystems is accounted for by additional terms in the core Hamiltonian. The Fock matrix for the active system embedded in the environment system is [31]

$$\mathbf{F}_{\text{emb}}^{\text{act}} = \mathbf{h}_{\text{core}} + \mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}} + \mathbf{g}(\gamma_{\text{emb}}^{\text{act}})$$
$$= \mathbf{h}_{\text{emb}} + \mathbf{g}(\gamma_{\text{emb}}^{\text{act}}), \tag{15}$$

where

$$\mathbf{V}_{\text{emb}} = \mathbf{g}(\gamma^{\text{act}} + \gamma^{\text{env}}) - \mathbf{g}(\gamma^{\text{act}}), \tag{16a}$$

$$\mathbf{h}_{\text{emb}} = \mathbf{h}_{\text{core}} + \mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}}. \tag{16b}$$

The embedding potential $\mathbf{V}_{\text{emb}}$ describes all the interactions (nonadditive part) between the active and environment subsystems [34]. Due to the subsystem densities [Eq. (12)] being constructed from disjoint subsets of orthogonal orbitals, the normally difficult-to-evaluate nonadditive kinetic potential (NAKP) terms [35] are exactly zero [31,34,36].

$\mathbf{P}_{\text{proj}}^{\text{env}}$ is a projection operator that enforces intersubsystem (orbital) orthogonality. There are different ways to define this

operator and we consider two in this work. The first definition was proposed by Manby *et al.* [31]. They use a parameter (μ) to shift the orbital energies of the environment to high energies—effectively meaning they will never be occupied. This projector is defined as

$$
\begin{aligned}
(\mathbf{P}_\mu^{\text{env}})_{ij} &= \mu\langle\psi_i^{\text{LMO}}|\mathbf{P}^{\text{env}}|\psi_j^{\text{LMO}}\rangle \\
&= \mu[\mathbf{S}\gamma^{\text{env}}\mathbf{S}]_{ij},
\end{aligned}
\tag{17}
$$

where μ is some large integer and $\mathbf{S}$ is the AO overlap matrix. $\mathbf{P}^{\text{env}}$ is a projector defined as

$$
\mathbf{P}^{\text{env}} = \sum_{l\in\mathcal{L}}|\psi_l^{\text{LMO}}\rangle\langle\psi_l^{\text{LMO}}|.
\tag{18}
$$

Here we use the notation $l \in \mathcal{L}$ to mean the sum over the set of occupied MO indices for the environment orbitals. It has been shown that μ is numerically robust and can usually be set to $\mu = 10^6$ [31,33]. In the limit that $\mu \to \infty$ this method is exact, as the environment orbitals will be pushed to infinite energy and thus will never be occupied. The action of this operator with the Fock matrix is

$$
(\mathbf{F} + \mathbf{P}_\mu^{\text{env}})|\psi_k^{\text{LMO}}\rangle = \epsilon_k^{\text{act}}|\psi_k^{\text{LMO}}\rangle,
\tag{19a}
$$

$$
\begin{aligned}
(\mathbf{F} + \mathbf{P}_\mu^{\text{env}})|\psi_l^{\text{LMO}}\rangle &= (\epsilon_l^{\text{env}} + \mu)|\psi_l^{\text{LMO}}\rangle \\
&\approx +\mu|\psi_l^{\text{LMO}}\rangle.
\end{aligned}
\tag{19b}
$$

Again, $k \in \mathcal{K}$ and $l \in \mathcal{L}$ are occupied LMOs of the active and environment subsystems, respectively. Qualitatively the orbital energies of the active system are left unchanged and the orbitals for the environment are pushed to very high energies as $\mu \gg \epsilon_l^{\text{env}}$—effectively suppressing transitions to these states and stopping hybridization.

The second approach, proposed by Kallay *et al.* [37], is to use the Huzinaga projector [38,39]:

$$
\begin{aligned}
\mathbf{P}_{\text{Huz}}^{\text{env}} &= -(\mathbf{F}\mathbf{P}^{\text{env}} + \mathbf{P}^{\text{env}}\mathbf{F}) \\
&= -\tfrac{1}{2}(\mathbf{F}\gamma^{\text{env}}\mathbf{S} + \mathbf{S}\gamma^{\text{env}}\mathbf{F}).
\end{aligned}
\tag{20}
$$

Note that the $-\tfrac{1}{2}$ prefactor is needed for closed-shell systems. This operator enforces orthogonality of the occupied orbitals of each subsystem [40]. The form of this operator increases the orbital energy for the occupied environment orbitals and leaves the active system unchanged. We write its action formally as

$$
(\mathbf{F} + \mathbf{P}_{\text{Huz}}^{\text{env}})|\psi_k^{\text{LMO}}\rangle = \epsilon_k^{\text{act}}|\psi_k^{\text{LMO}}\rangle,
\tag{21a}
$$

$$
\begin{aligned}
(\mathbf{F} + \mathbf{P}_{\text{Huz}}^{\text{env}})|\psi_l^{\text{LMO}}\rangle &= (\epsilon_l^{\text{env}} - 2\epsilon_l^{\text{env}})|\psi_l^{\text{LMO}}\rangle \\
&= -1\epsilon_l^{\text{env}}|\psi_l^{\text{LMO}}\rangle.
\end{aligned}
\tag{21b}
$$

It is usually the case that for the occupied molecular orbitals ϵ_l^{env} is negative and so Eq. (21b) shows how the environment ϵ_l^{env} orbitals are normally shifted to positive energies and thus will not be filled. If any occupied MO has a positive ϵ_l^{env}, then Eq. (20) can be modified to treat such systems correctly; this operator is known as the "Fermi-shifted Huzinaga operator" [41]. Both the μ shift and Huzinaga projectors work by the same logic, increasing the energy level of environment orbitals to suppress transitions from the active region. In practice they therefore give similar results. However, the Huzinaga

formalism guarantees that $[\mathbf{P}_{\text{Huz}}^{\text{env}}, \mathbf{F}_{\text{emb}}^{\text{act}}] = 0$ and removes the parameter needed by the μ-shift method [42].

The energy of the active system embedded in the environment is given by

$$
\begin{aligned}
E[\gamma_{\text{emb}}^{\text{act}}; \gamma^{\text{act}}, \gamma^{\text{env}}] ={}& \mathcal{E}[\gamma_{\text{emb}}^{\text{act}}] + E[\gamma_{\text{env}}] + \mathbf{g}(\gamma^{\text{act}}, \gamma^{\text{env}}) \\
&+ \text{Tr}\big((\gamma_{\text{emb}}^{\text{act}} - \gamma^{\text{act}})(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}})\big),
\end{aligned}
\tag{22}
$$

colloquially denoted as a DFT-in-DFT calculation.

We use the same notation as Claudino and Mayhall [33], where $\mathcal{E}$ differs from E as it allows for different functionals to be applied and is computed from the embedded density matrix of the active system. Note that Eq. (15) is solved self-consistently to give $\gamma_{\text{emb}}^{\text{act}}$. Equation (22) reduces to Eq. (13) for the case that the active and environment regions are treated at the same level of theory [33].

Importantly $\mathcal{E}[\gamma_{\text{emb}}^{\text{act}}] = \text{Tr}(\gamma_{\text{emb}}^{\text{act}}\mathbf{h}_{\text{core}}) + \mathbf{g}(\gamma_{\text{emb}}^{\text{act}})$ and does not involve $\mathbf{V}_{\text{emb}}$ or $\mathbf{P}_{\text{proj}}^{\text{env}}$. The final term in Eq. (22) is a first-order correction that accounts for the difference between $\mathbf{g}(\gamma^{\text{act}}, \gamma^{\text{env}})$ and $\mathbf{g}(\gamma_{\text{emb}}^{\text{act}}, \gamma^{\text{env}})$, and corrects for the fact that in general $\gamma^{\text{act}} \neq \gamma_{\text{emb}}^{\text{act}}$ [43].

This projection-based embedding approach then allows for the active system to be treated using some wave-function level of theory and therefore studied using a quantum computer. The electronic energy for this is given by [33]

$$
\begin{aligned}
E[\Psi_{\text{emb}}^{\text{act}}; \gamma^{\text{act}}, \gamma^{\text{env}}] ={}& \langle\Psi_{\text{emb}}^{\text{act}}|\mathbf{H}_{\text{emb}}|\Psi_{\text{emb}}^{\text{act}}\rangle + E[\gamma_{\text{env}}] \\
&+ \mathbf{g}(\gamma^{\text{act}}, \gamma^{\text{env}}) - \text{Tr}(\gamma^{\text{act}}(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}})).
\end{aligned}
\tag{23}
$$

Importantly $\mathbf{H}_{\text{emb}} = \mathbf{h}_{\text{emb}} + \mathbf{g}(\Psi_{\text{emb}}^{\text{act}})$, where $\mathbf{g}(\Psi_{\text{emb}}^{\text{act}})$ is the two-electron operator for a given wave-function method and $\mathbf{h}_{\text{emb}}$ is the embedded core Hamiltonian [Eq. (16b)] which depends on γ^{act} and γ^{env} [44]. As the embedding terms have been included in $\mathbf{H}_{\text{emb}}$, the final correction term is therefore slightly different from Eq. (22) [43]. The wave-function calculation in Eq. (23) includes contributions from $(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}})$, similar to $\text{Tr}(\gamma_{\text{emb}}^{\text{act}}(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}}))$. The correction therefore only requires subtracting $\text{Tr}(\gamma^{\text{act}}(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}}))$, unlike in Eq. (22), where $\mathcal{E}$ does not use $(\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}})$ to calculate the energy of the active system.

B. MO localization methods

In this work, we only use the subsystem projected AO decomposition (SPADE) and intrinsic bonding orbitals (IBOs) localized molecular orbitals [33,45]. The motivation for using SPADE is primarily that it does not require a parametrized heuristic to determine the active and environment subsystems. Furthermore, there are two main benefits of the SPADE approach. First, the singular values are analogous to occupation numbers. This is due to the similarity with a natural population analysis, which has been shown to avoid some of the flaws of the Mulliken analysis [33,46]. Second, the SPADE orbital construction involves a projector onto the set of orthogonal AOs associated with the atoms in the active subsystem. This differs from most localization schemes that attempt to localize orbitals in atoms or bonds, whereas SPADE orbitals are local only in the sense that they remain in their native subsystems,

which is one of the requirements for successful embedding [33].

IBOs were used as they only depend on the intrinsic atomic orbital charges, rather than Mulliken charges which change erratically depending on the basis set used [45]. IBOs are therefore always well defined, whereas other localization methods—such as Pipek-Mezey orbitals [47], which depend on the Mulliken charges [48]—are unphysically tied to the basis set used [45]. For the IBO approach, we calculate the percentage of the ith LMO over atoms a user defines as the active subsystem. Any LMO that has a percentage higher than a threshold (here 95%) was assigned to the active region. The SPADE approach does not require this predefined threshold hyperparameter. However, it does use a function of the molecular orbital coefficient matrix. In practice, these methods give similar active orbitals when a single atom is selected as the active region; however, SPADE tends to include a greater number of orbitals when multiple atoms are active. Additionally, increasing the number of active atoms does not necessarily increase the size of the active space found by SPADE. Further details on both approaches are given in Appendix A.

The effects of different localization methods for this embedding method, such as Pipek-Mezey (PM) [47], Foster-Boys (FB) [49], Edmiston-Ruedenberg (ER) [50] and fourth moment (FM) localization [51], would be an interesting area to explore. Our software package Nbed can run any method given by PySCF, and users can also build their own localization strategies themselves [52].

III. METHOD

We studied the performance of our wave-function projection-based embedding method on a selected set of molecular systems. We have developed a PYTHON package, Nbed, that utilizes the PySCF and Openfermion quantum chemistry packages to build each embedded model [53,54]. The package outputs a qubit Hamiltonian for the wave-function portion of an embedded problem and the classical energy corrections from density functional theory. This is freely available for use on GitHub [52].

For all calculations presented, the minimal STO-3G basis set was employed. Each global DFT calculation performed, prior to orbital localization, used the B3LYP5 functional. The IBO or SPADE localization procedures are used in order to isolate the molecular orbitals to the active and environment subsystems from preselected active atoms [33,45]. For the IBO procedure, in order to assign the active and environment molecular orbitals, we calculate the percentage of the ith LMO over atoms a user defines as the active subsystem. Any LMO that has a percentage higher than a predefined threshold (we used 95%) is assigned to the active region. Appendix A goes into further detail on each localization strategy. We performed both the μ-shift and Huzinaga methods for each. A Hartree-Fock calculation for the active system, using the modified core Hamiltonian [Eq. (16b)], was performed for each molecular system. The second quantized molecular Hamiltonian was then constructed with Openfermion and converted to a qubit Hamiltonian using the in-built Jordan-Wigner transformation [55]. Post-Hartree-Fock (HF) methods were performed with

PySCF. The frozen core approximation was not used and all virtual orbitals were included in the wave-function calculations. Only the occupied environment molecular orbitals were removed from the wave-function calculations of the active systems. To achieve this, the columns of C (the matrix of MO coefficients) associated with the environment were masked and not considered by further post-HF methods on the embedded active system (the WF part of a WF-in-DFT calculation). The removal of these (occupied) environment orbitals is what gives a qubit reduction (when constructing the second quantized molecular Hamiltonian). We note that this removal approach is slightly different from the implementation of Chulhai and Goodpaster [42], where these orbitals are left in the embedded calculation. Our approach is justified as the environment MOs have been projected out of the SCF problem. More qualitatively, in Chulhai and Goodpaster's approach the results of the Huzinaga and μ-shifted approaches are very similar [42]. For the μ-shifted approach, the occupied environment orbitals are shifted to such high energies that they remain unoccupied in subsequent post-HF calculations. This is not the case for the Huzinaga method but, as it gives similar results to the μ-shifted technique, removing the associated (occupied environment) orbitals of the Huzinaga method falls in line with them not being able to be occupied in the μ-shifted approach.

For the single-point electronic structure calculations we perform a CCSD-in-DFT calculation (active subsystem treated at the CCSD level). Each result is then compared to full-system CCSD(T) calculations. Each molecular geometry was obtained from PubChem [56]. The potential energy surface of an OH bond stretching in water was calculated using FCI-in-DFT, where the embedded molecular Hamiltonian at each geometry was diagonalized to find the ground-state energy of the active system. This was compared to a full-system FCI calculation at each step. As the PBE model requires a full-system DFT calculation to determine the active and environment subsystems, we also report these DFT results. Whilst we have applied this method here to the OH functional group in water, we note that this projection-based embedding method generalizes to other functional groups and molecular fragments as seen in Refs. [33,41,42].

IV. RESULTS AND DISCUSSION

In the following sections, we apply our model to different molecular systems.

A. Molecular ground-state energy

In order to assess the ability of the embedding procedure, we selected a test bed of molecular structures, which are summarized in Fig. 1. The active atoms considered at a more expensive level of theory are highlighted in green. The choice of these molecules was motivated by selecting compounds commonly encountered by chemists. To date, most quantum computing studies consider only the smallest molecular systems (often H_2, LiH, and BeH_2) [17,58,59], due to current quantum computing constraints—low numbers of qubits and high error rates. The goal of this paper is to show that this embedding approach will allow larger systems to be studied
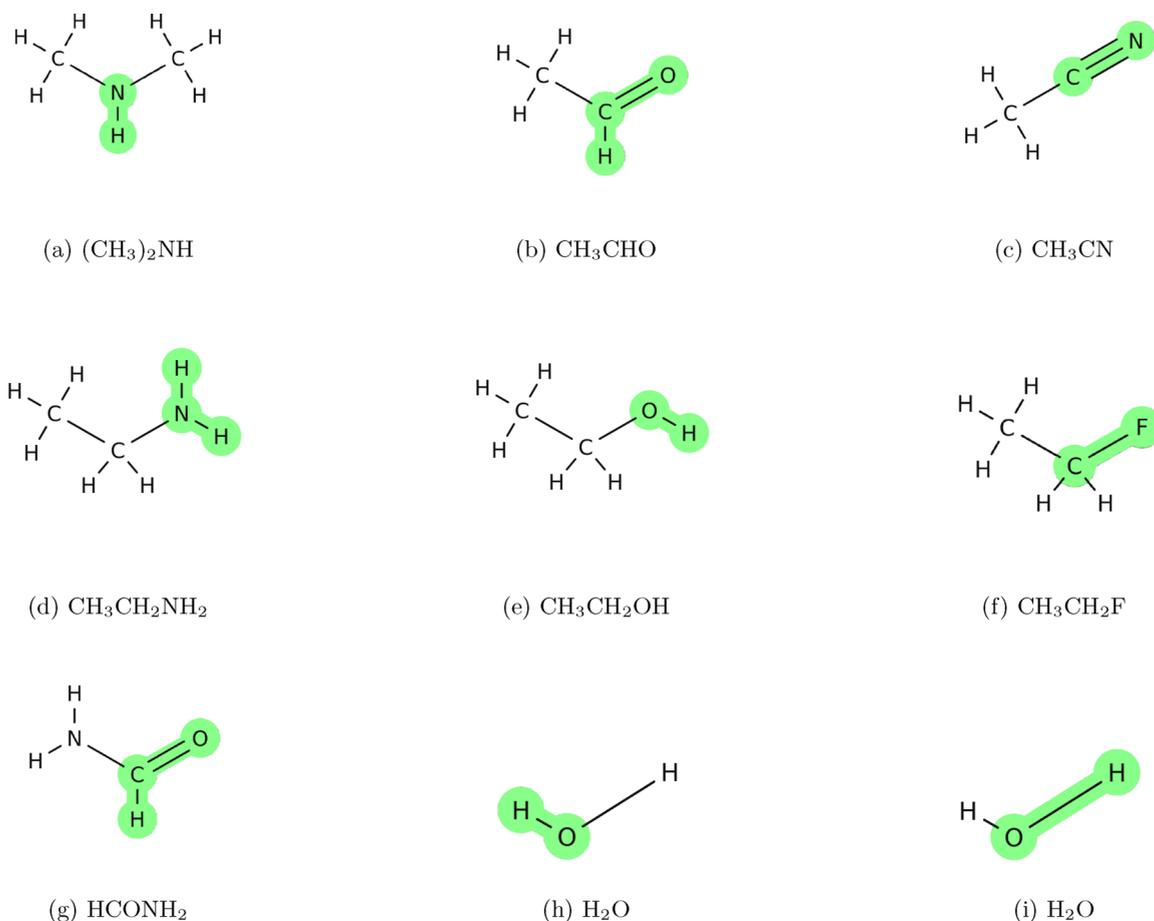
FIG. 1. Planar representations of the molecules used in embedding calculations. Atoms shaded in green were selected as active for localization procedures. Images were generated using MolView [57]. (a) N-methylmethanamine, (b) acetaldehyde, (c) acetonitrile, (d) ethanamine, (e) ethanol, (f) fluoroethane, (g) formamide, (h) water (fixed bond active), and (i) water (stretching bond active).

on such devices. Figure 2 reports the results for SPADE localized CCSD-in-DFT embedding molecular ground state energy calculations for the molecules in Fig. 1. Numerical values of these results are available in Appendix D along with results for the same calculations using IBO localized orbitals.

The embedded FCI Hamiltonians (describing the active region) output using both localization methods was significantly reduced in number of terms and qubit counts compared to the full-system FCI Hamiltonian. However, they still exceed the limit of what is practical to exactly solve using classical computers. We therefore performed classical CCSD-in-DFT calculations, the results of which are given in Fig. 1(a). Our results show increased accuracy in CCSD-in-DFT calculated molecular ground-state energies, compared to full-system DFT. We benchmarked these approaches compared to full-system CCSD(T), as full-system FCI was not possible. The metric $|\Delta E| = |E_{\mathrm{expt}} - E_{\mathrm{CCSD(T)}}|$ was used to approximate the true error $|\Delta E_{\mathrm{true}}| = |E_{\mathrm{expt}} - E_{\mathrm{FCI}}|$, where E_{expt} is the ground-state energy calculated via different procedures as specified and $E_{\mathrm{CCSD(T)}}$ is a full-system CCSD(T) reference ground-state energy that is used to approximate each (full-system) FCI ground-state energy E_{FCI}. Our results show that PBE embedding gives ground-state energies closer to the reference value, namely, the full-system CCSD(T) energy. It is known that different DFT functionals will give different ground-state

densities and thus energies [60]; however, there will always be a true ground-state wave function and energy. The results in Fig. 2 show that we can improve on the results of a DFT calculation by getting closer to the ground-state energy; in our case we approximate with CCSD(T). Qualitatively, we attribute the improvement of our PBE method to it including different correlation effects not captured by the DFT calculation. Further evidence of this will be seen in the next section on strong correlation.

Typically, results for the μ-shift and Huzinaga projectors are very similar. This is expected as the active and environment subsystems were described in the supermolecular basis in our implementation of PBE [37,42]. If the subsystems were described in an alternate basis, it has been shown that the Huzinaga operator outperforms the μ-shift approach [42]. In our results, the Huzinaga projector usually produces marginally more accurate energies compared to the μ-shifted implementation. We attribute this to the Huzinaga approach being based on a formally exact embedding, while the μ-shift embedding is approximate due to a finite shift value being used.

The number of qubits describing the embedded FCI Hamiltonian (of a FCI-in-DFT problem) will be the same between the embedding methods [Fig. 1(b)]. This is due to the number of qubits depending on how many spin orbitals are considered in the embedded active system. In this work, only the occupied

Spade Localization

(a) Molecular Ground-state Energy.

(b) Number of qubits in Hamiltonian.

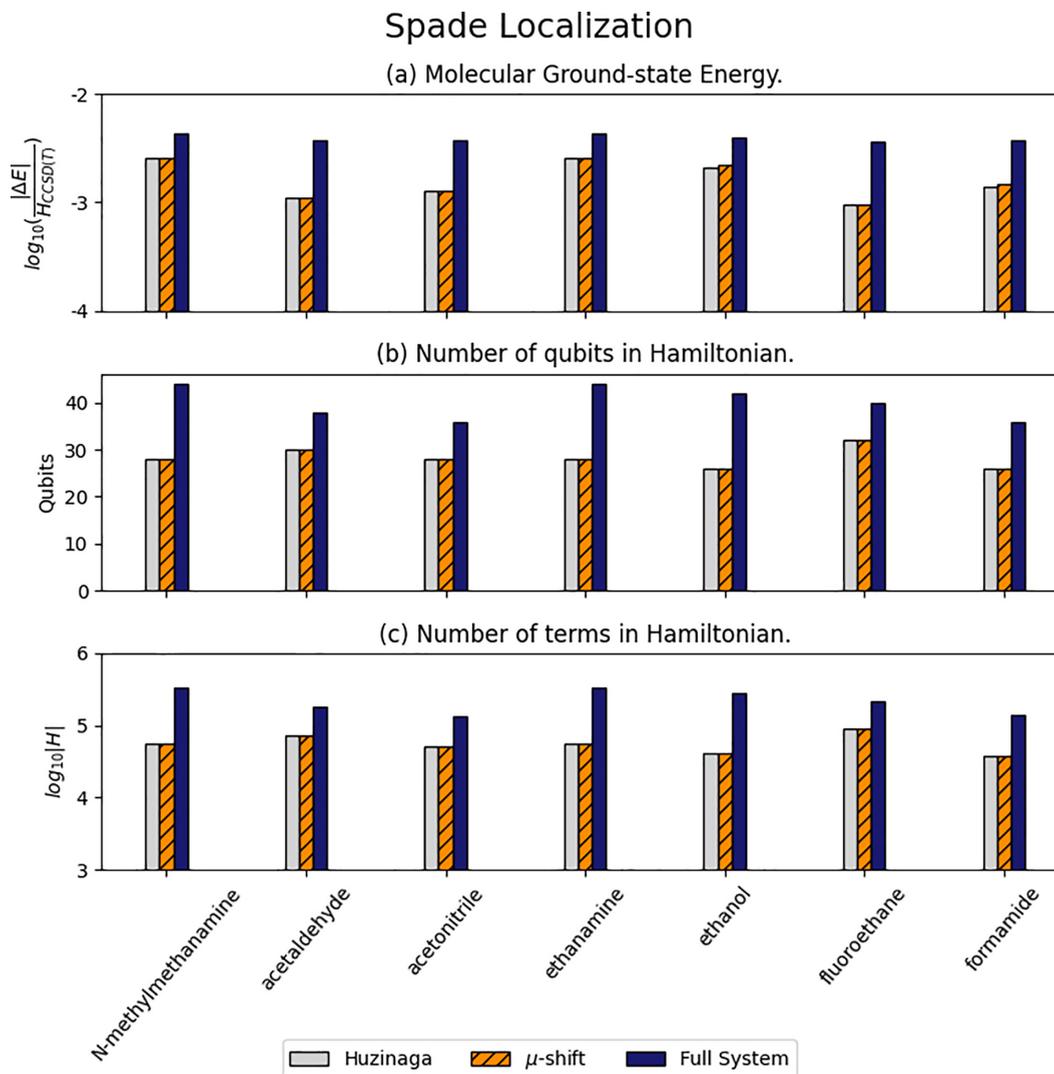### (c) Number of terms in Hamiltonian.

FIG. 2. Results for embedding of small molecules (Fig. 1) using the SPADE localization method. Bar chart (a) reports the ground-state energy error for small molecules compared to full-system CCSD(T): $|\Delta E| = |E_{\text{expt}} - E_{\text{CCSD(T)}}|$. For the blue result E_{expt} is the full-system DFT (B3LYP) ground-state energy, for the orange result E_{expt} gives the μ-shift CCSD-in-DFT embedding energy, and for the grey result E_{expt} is the Huzinaga CCSD-in-DFT embedding energy. (b) The number of qubits needed to describe the embedded Hamiltonian, with the reference showing the number required for the full-system FCI Hamiltonian. (c) The number of terms in the Jordan-Wigner encoded embedded FCI-in-DFT Hamiltonian for each molecule with the blue bar representing the number of terms in the FCI Hamiltonian of the full system.

molecular orbitals of the environment are removed from the embedded active subsystem WF calculations (which leads to a qubit reduction). To further reduce the qubit count, virtual (unoccupied) molecular orbitals should also be removed from the embedded active subsystem calculations. This requires partitioning of the virtual space between the active and environment subsystems. Recent work by Yuan and co-workers shows that truncating the virtual space can still give reliable estimates of both energies and molecular properties and we anticipate that this result will also be found if virtual environment orbitals are removed from the active embedded subsystem [61]. We leave this to future work, but note that it could lead to a significant further reduction in the number of qubits.

The number of terms in the Jordan-Wigner encoded qubit FCI Hamiltonian of the embedded WF problem (FCI-in-DFT

problem), $|H|$, is typically very similar between the two projection methods, as shown in Fig. 1(c). This is expected as the number of molecular orbitals used to describe the embedded problem is the same between the different projection methods.

In comparing the two localization methods, we find that for acetonitrile and formamide, SPADE and IBO partition the active system in a similar way. This results in a similar number of active MOs and hence the ground-state estimation and resource requirements are very similar for these systems. For the majority of the molecules we study, SPADE includes more MOs, resulting in significantly more accurate ground-state energies while still reducing the size of the Hamiltonian. However, by reducing the threshold for assigning the localized MOs from IBO to the active region, additional MOs could be included, giving a similar result. See Appendix D for further details.
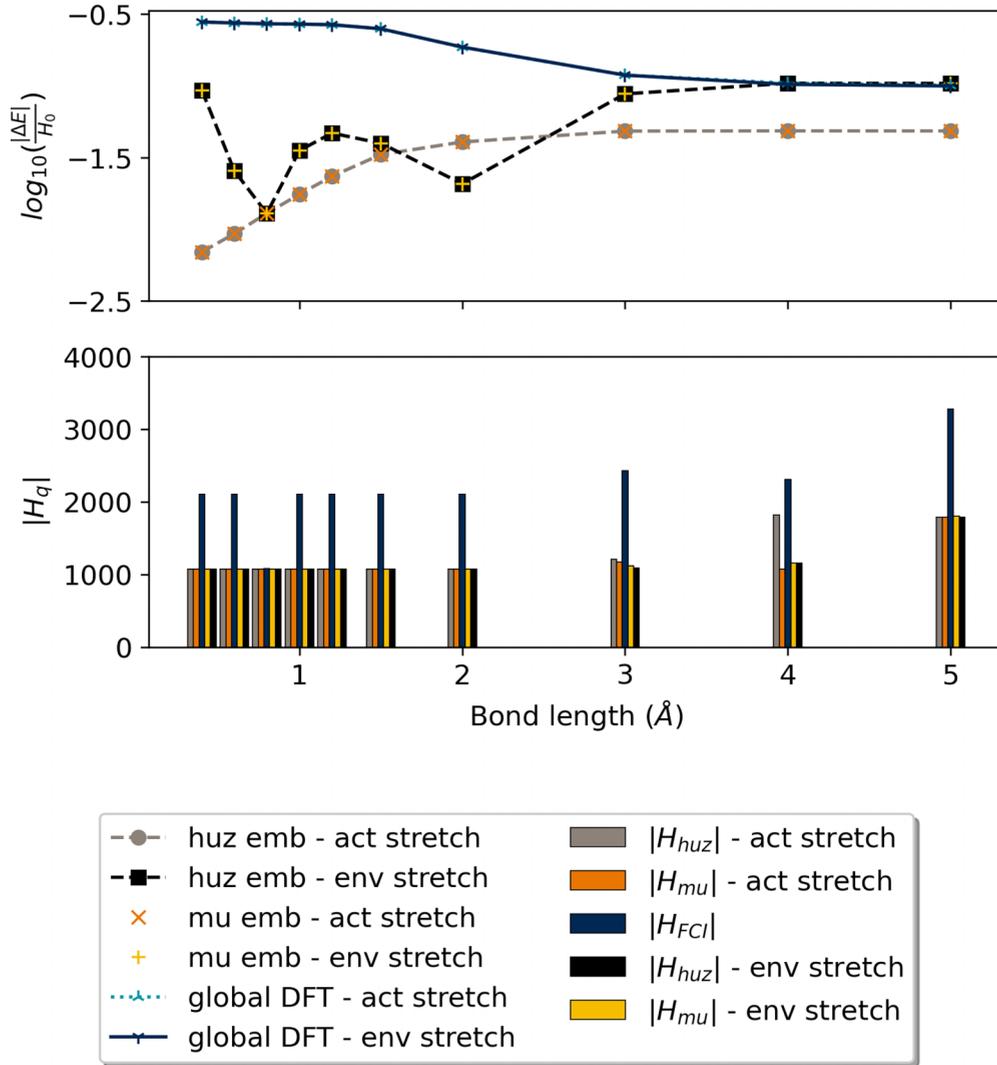
FIG. 3. Potential energy curve for H_2O, with changing OH bond length. Active (act) stretch result has the changing OH bond as the active region and environment (env) stretch result has the fixed OH bond selected as the active region. These results use SPADE localization. For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being four in all cases. The top plot reports the $\log_{10}$ error with respect to the exact FCI ground-state energy (E_{FCI}) of the whole system, where $|\Delta E| = |E_{expt} - E_{FCI}|$. Here E_{expt} is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full-system Hamiltonian, the orange and yellow results are for μ-shifted embedded Hamiltonians, while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are available in Appendix E.

B. Strong correlation

The impact of active region selection is demonstrated by our results shown in Fig. 3, where SPADE localization was used in the embedding calculations (results for IBO localization are provided in Appendix E). We consider the bond dissociation of an OH bond in water, where at high bond lengths, a correlated state is created [62,63]. We perform projection-based FCI-in-DFT calculations, at different molecular geometries, for two different active regions. One has the atoms in the fixed OH bond set active and the other has the atoms in the changing OH bond set active. We show this pictorially in Figs. 1(h) and 1(i). These results are compared to full-system FCI calculations. At near-equilibrium bond lengths, we see a similar performance between the different active systems (Fig. 3). This is due to the symmetrical

structure of H_2O; hence at low bond lengths there is little difference between the two active regions. In fact, the third data point gives results for the scenario where both OH bonds are the same length and consequently is why the results for the different active regions are the same here. However, in the correlated regime—at large bond lengths—selecting the active region to encompass the stretched atoms leads to significant improvements in energy calculation over DFT alone. This is due to correlation being effectively captured in the wave-function calculation. In contrast, the full DFT calculation is plagued by deficiencies of current approximate exchange-correlation functionals [64,65]. We see in Fig. 3 that the global DFT calculation overestimates the bond dissociation energy. This problem is attributed to static correlation [64]. As there is no systematic way to improve the approximate

exchange-correlation functionals, the way forward to describe such systems may be hybrid quantum-classical embedding. Here quantum processors could be exploited most effectively by application to only those regions of a molecule that are highly correlated.

These results also show the importance of active region selection in an embedded wavefunction calculation. There are two main factors that need to be considered. The first is that the nonadditive two-electron energy, although in principle exact with the true exchange-correlation functional, is approximated and thus leads to missing off-diagonal information resulting in errors [5]. The active region must therefore be chosen to minimize this where possible, which in practice can be hard to do. Second, a highly correlated active region would benefit the most from a higher-level wave-function calculation, be it FCI, CCSD, or a quantum algorithm such as VQE or QPE. It is well known that static correlation effects are present during bond dissociation and required to describe certain transition-metal species [66–69]. To formalize a more mathematical way to determine when static correlation occurs, we recommend using common diagnostic tools such as the T_1 [70] and D_1 [71] measures. A review of these and alternate techniques can be found in Ref. [72]. These techniques can be used to select an active region in a more informed manner.

Finally, we note that all the DFT calculations performed in this work were restricted. At large bond lengths, it is known that spin-unrestricted calculations can often better describe bond dissociation; however, the major drawback of the method is spurious symmetry breaking [73]. Bearing this in mind, our results show that WF embedding combined with a restricted calculation can still capture major correlation effects.

V. CONCLUSIONS

For a small collection of molecules, too large to study completely (full system) on currently available quantum hardware, we have shown that the PBE method allows a smaller active system to be studied using less resources on a quantum computer and the calculated energies of such embedding calculations are closer to the "gold standard" CCSD(T) of the full system compared to full-system DFT. Furthermore, we have shown its ability to capture the effects of strong correlation by investigating the bond dissociation of H_2O.

We use the projection-based embedding technique [31] to reduce the size of an electronic structure calculation studied at the wave-function level. The molecular problem is split into active and environment parts, each solved using different levels of theory. The active part is treated using a wave-function approach and an embedded qubit Hamiltonian is generated. Solving this provides $E_{act}^{WF} = \langle \Psi_{emb}^{act} | \mathbf{H}_{emb} | \Psi_{emb}^{act} \rangle$. The whole system and environment are treated using density functional theory and the overall electronic energy is found via an additive procedure [42–44]. This is similar to the ONIOM (which stands for "our own n-layered integrated molecular orbital and molecular mechanics") subtractive framework [74]. What is included in the active region can be modified and thus the size of the quantum problem varied. This allows users to tune their
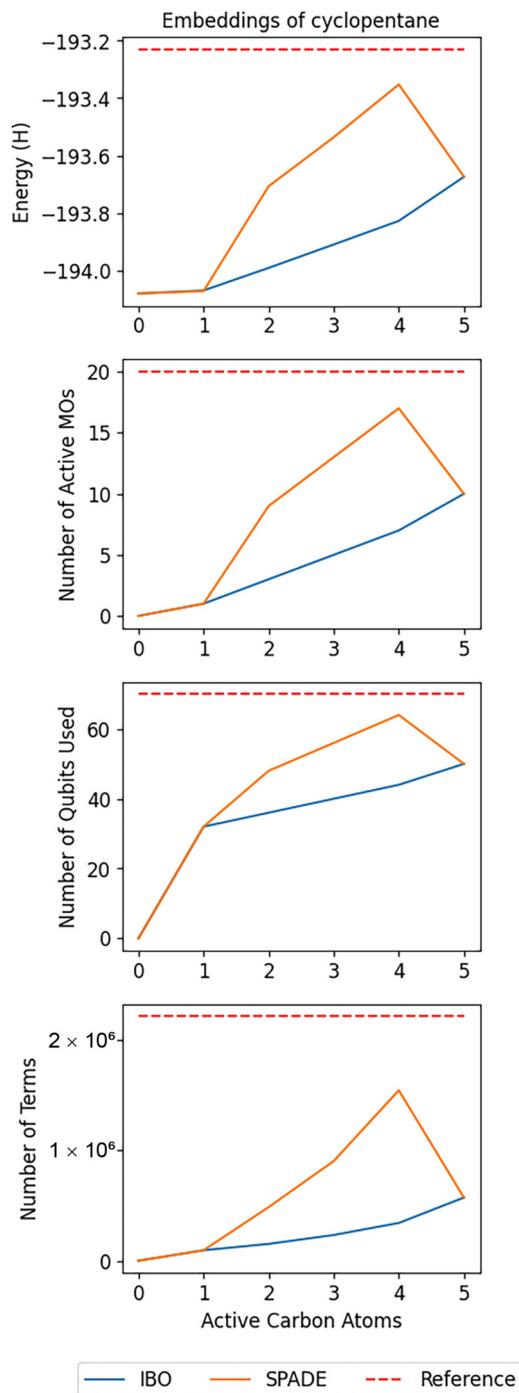

FIG. 4. Results of cyclopentane embedding, with increasing numbers of carbon atoms assigned to the active region. For each, IBO localization is shown in blue and SPADE in orange; results for the μ-shift and Huzinaga projectors are overlapping and indistinguishable. Different projectors do not give noticeably distinct results. (a) Ground-state energy. (b) Number of occupied molecular orbitals assigned to the active region. (c) Number of qubits in the output Hamiltonian. (d) Number of terms in the Jordan-Wigner encoded qubit Hamiltonian.

problem to available hardware and so it is possible to simulate large molecular problems on small quantum devices.
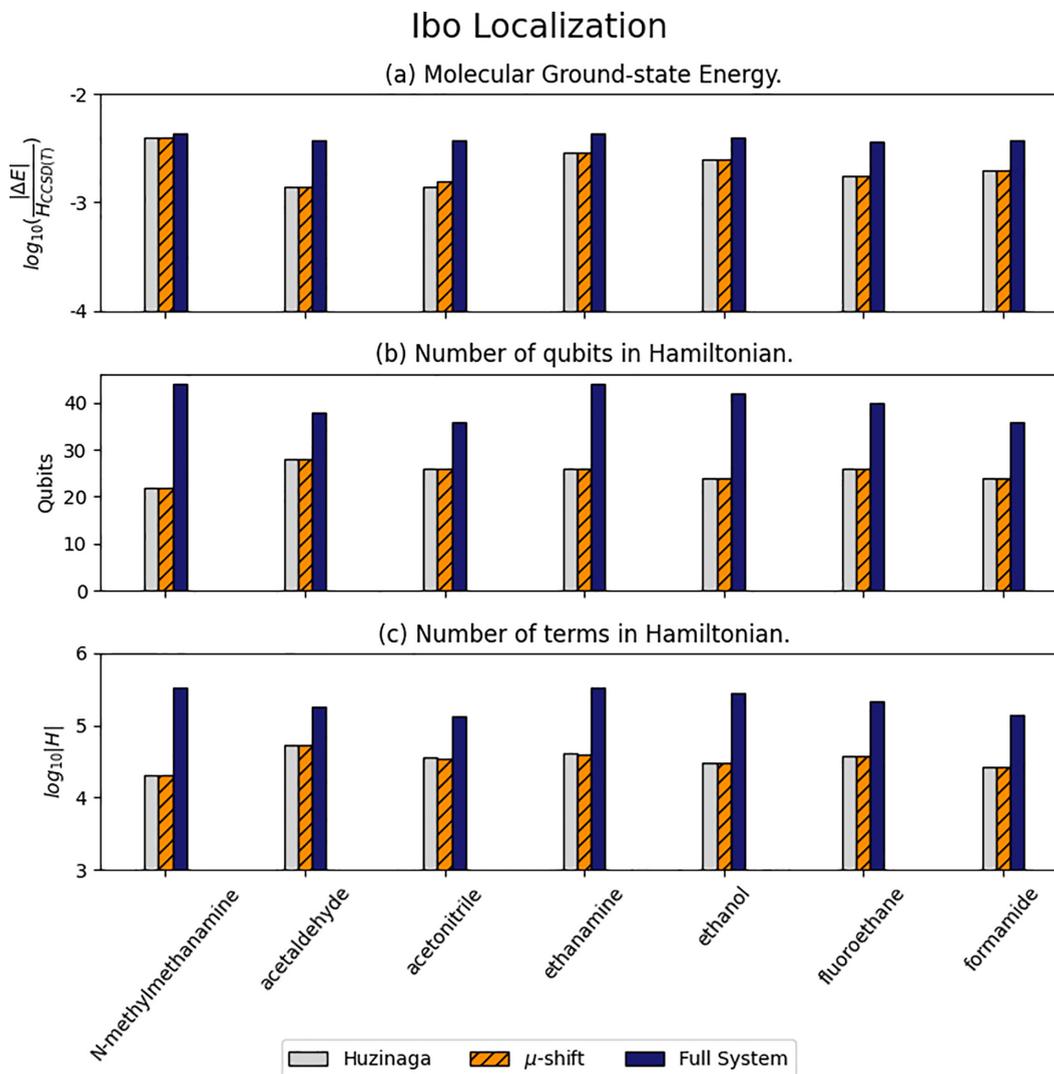
Ibo Localization

FIG. 5. Results for embedding of small molecules using the IBO localization method. (a) Ground-state energies error for small molecules, with full-system DFT energy error as reference, μ-shift CCSD-in-DFT embedding energy in orange, and Huzinaga CCSD-in-DFT embedding in grey. All error values are calculated with respect to whole-system CCSD(T) energies. (b) The number of qubits required to describe the embedded FCI-in-DFT Hamiltonians, with reference showing the number required for the full-system FCI Hamiltonian. (c) The number of terms in the Jordan-Wigner encoded FCI-in-DFT qubit Hamiltonian for each molecule. Again the reference gives the number needed for the full-system FCI Hamiltonian.

As this approach generates an embedded qubit Hamiltonian, it is agnostic to the quantum algorithm used to solve $\mathbf{H}_{\mathrm{emb}}$. NISQ-friendly approaches such as the VQE algorithm can therefore be used, but also fault-tolerant methods such as QPE [12,17].

Moreover, as our method outputs a qubit Hamiltonian, different resource reduction techniques can be used in conjunction with it; for example, the contextual-subspace approach of Kirby and Love [75], or the entanglement forging approach of Eddins *et al.* [63]. Similarly, the $\mathbb{Z}_2$ symmetries of the problem can also be removed via qubit tapering [76]. Approaches to improve near-term NISQ calculations can also be used (for example, see Refs. [77–79]).

As our method does not rely on imposing constraints on the system studied or costly parameter fitting, it may be reasonably combined with other hybridization techniques which do [20,80].

Further work is planned to develop this method. As significant resource reduction is achieved by localization of only the occupied orbitals, virtual orbital localization could lead to a greater reduction in computational resources [81]. In the context of this work, if virtual LMOs are included in the active and environment subsystems, respectively, then the number of qubits will reduce by how many are included in the environment. This will also have the effect of decreasing the total number of Pauli operators in the associated embedded qubit Hamiltonian.

We anticipate that our code will allow researchers to study molecules of real chemical interest on quantum computers. We welcome readers to make use of this, which is freely available on GitHub [52].
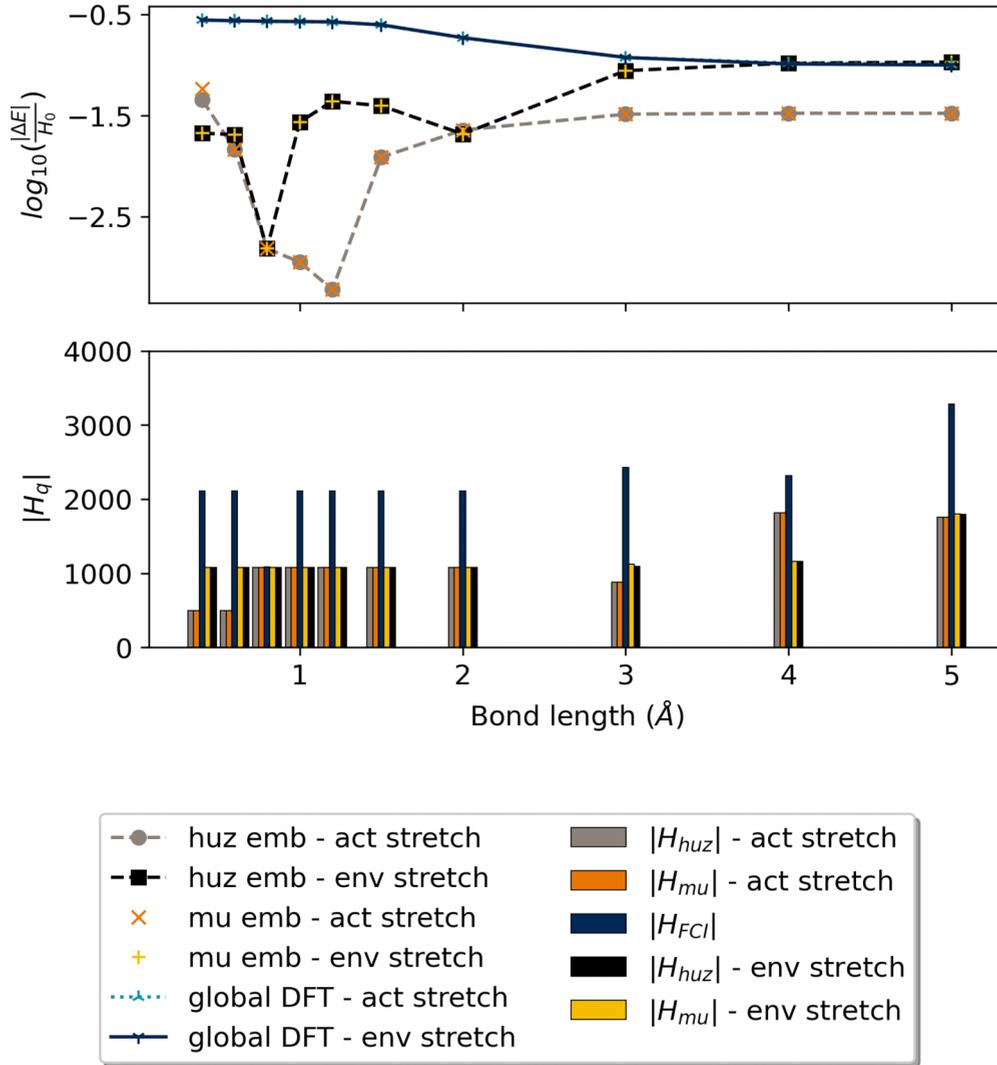
FIG. 6. Potential energy curve for H_2O, with changing OH bond length. Active (act) stretch result has the changing OH bond as the active region and environment (env) stretch result has the fixed OH bond selected as the active region. These results use IBO localization (95% threshold). For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being three or four in all cases. The top plot reports the $\log_{10}$ error with respect to the exact FCI ground-state energy (E_{FCI}) of the whole system, where $|\Delta E| = |E_{expt} - E_{FCI}|$. Here E_{expt} is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full-system Hamiltonian, the orange and yellow results are for μ-shifted embedded Hamiltonians, while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are provided in Tables VII and VIII.

APPENDIX A: LOCALIZATION

In this section, we describe how to generate a set of localized molecular orbitals (LMOs) from the canonical molecular orbitals (CMOs) of a global DFT calculation. This step is required in order to define each subsystem according to a certain metric discussed in further detail here.

We employ two methods to generate a set of localized molecular orbitals: the intrinsic bond orbitals (IBOs) [45] and subsystem projected AO decomposition (SPADE) [31]. Other methods can be used, such as Pipek-Mezey (PM) [47], Foster-Boys (FB) [49], Edmiston-Ruedenberg (ER) [50], and fourth-moment (FM) localization [51].
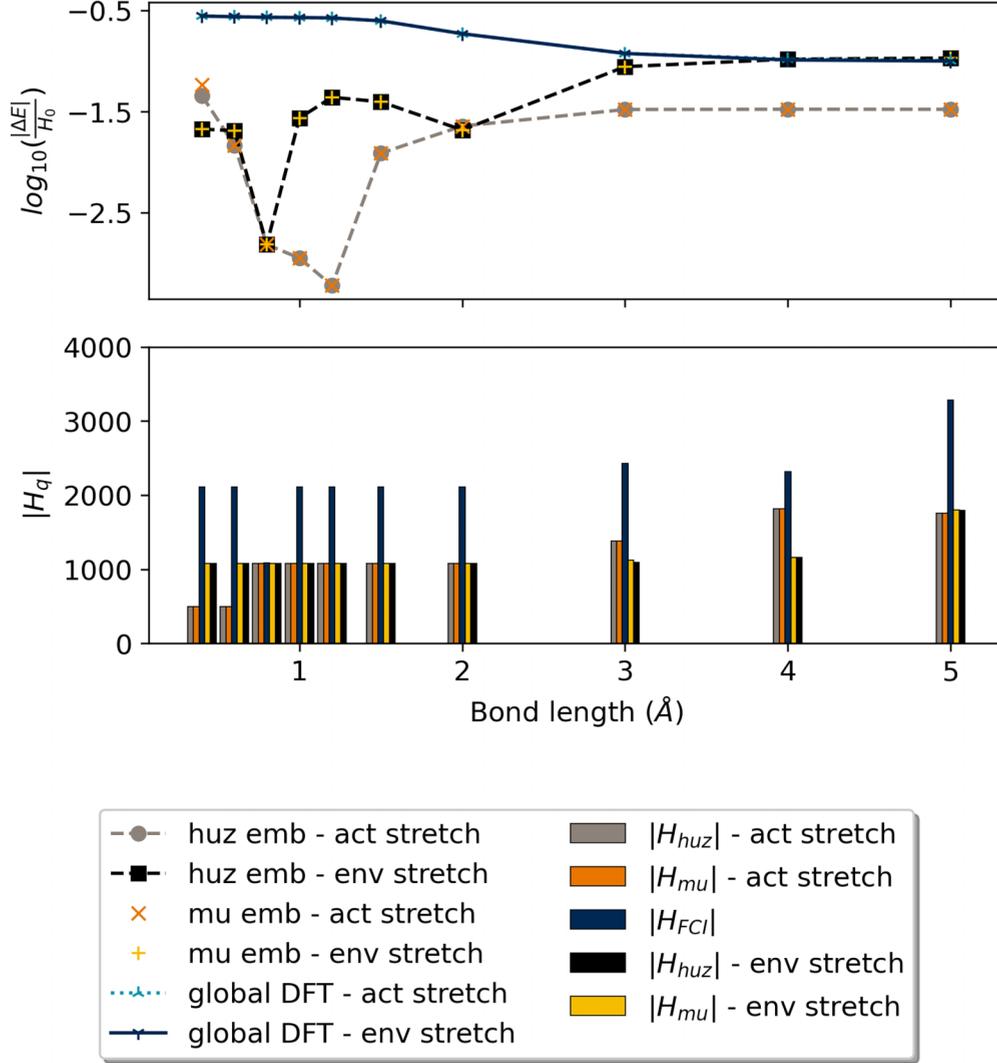
FIG. 7. Potential energy curve for H_2O, with changing OH bond length. Active (act) stretch result has the changing OH bond as the active region and environment (env) stretch result has the fixed OH bond selected as the active region. Results use IBO localization (see raw data in the Supplemental Material [84] for threshold values; lowest setting was 90%). For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being four in all cases. The top plot reports the $\log_{10}$ error with respect to the exact FCI ground-state energy (E_{FCI}) of the whole system, where $|\Delta E| = |E_{expt} - E_{FCI}|$. Here E_{expt} is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full-system Hamiltonian, the orange and yellow results are for μ-shifted embedded Hamiltonians, while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are provided in the Supplemental Material [84].

1. SPADE

SPADE is designed to localize electrons to the general region of an active site and environment [33]. This avoids the need for a threshold used to define the active and environment systems, as needed by the PM, FB, ER, FM, and IBO approaches. The next section goes into more details about this.

We first restrict the **C** matrix to only the occupied molecular orbitals $\{\psi_{occ}\} \subset \{\psi\}$. As our aim is to localize electrons to one region, virtual orbitals need not be localized. The remaining MOs are made mutually orthogonal by rotation with the matrix defining the overlap of atomic orbitals, **S** [33]:

$$\bar{\mathbf{C}} = \mathbf{S}^{0.5}\mathbf{C}_{occ}. \tag{A1}$$

We wish to find the relative contribution of the active region AOs to the occupied MOs, so we now restrict the matrix to only contributions from AOs associated to atoms in the active region $\bar{C}_A$. By making a singular value decomposition of this matrix [33],

$$\bar{C}_A = P_A\bar{\mathbf{C}}, \tag{A2}$$

$$\bar{C}_A = \mathbf{U}_A\boldsymbol{\Sigma}_A\mathbf{V}_A^\star, \tag{A3}$$

we can determine a basis which localizes the electrons of the active region. The column vectors of $\mathbf{V}_A$ are identical to the eigenvectors of $\bar{\mathbf{C}}_{occ}^{A\dagger}\bar{\mathbf{C}}_{occ}^A$ [33]. A new set of orbitals is then obtained through a rotation of the original MOs $\mathbf{C}_{occ}$ as

$$\mathbf{C}_{occ}^{SPADE} = \bar{\mathbf{C}}_{occ}\mathbf{V}_A, \tag{A4}$$

TABLE I. Full-system reference values for embedding calculations of small molecules, as shown in Figs. 2 and 5. ϵ_{DFT} gives the difference between full-system Restricted Kohn-Sham DFT, using the B3LYP functional, and CCSD(T). Q and $|H|$ give respectively the number of qubits and terms in the Jordan-Wigner encoded qubit Hamiltonian of the full system. All energies reported in hartree.

Molecule	ϵ_{DFT}	Q	$|H|$
N-methylmethanamine	0.573	44	338967
acetaldehyde	0.569	38	182702
acetonitrile	0.485	36	136075
ethanamine	0.573	44	329299
ethanol	0.609	42	283020
fluoroethane	0.637	40	217385
formamide	0.619	36	138231

where $\mathbf{C}_{\text{occ}}^{\text{SPADE}}$ are the SPADE orbitals. The singular values $\{\sigma\}$ of $\boldsymbol{\Sigma}_A$ allow appropriate partitioning of the orbital subspaces into active and environment subsystems [33]. The largest difference between successive singular values gives the partition of the SPADE MOs. Formally we write this as

$$m_{\text{occ}}^A = \max_i\{\sigma_i - \sigma_{i+1}\}, \tag{A5}$$

where m_{occ}^A is the index i where the largest difference occurs. The SPADE MOs (columns of $\mathbf{C}_{\text{occ}}^{\text{SPADE}}$) up to this index are the active MOs and the remaining orbitals (other columns of $\mathbf{C}_{\text{occ}}^{\text{SPADE}}$) are the environment orbitals.

2. Intrinsic bonding orbitals

IBOs are constructed as a linear combination of intrinsic atomic orbitals (IAOs), where the number of atoms the orbital charge is spread over is minimized.

We take the canonical molecular orbitals defined as

$$|\psi_i\rangle = \sum_j^K \mathbf{C}_{ji}|\phi_j\rangle, \tag{A6}$$

where $|\phi_j\rangle \in B_1$ are basis functions from a basis set B_1. A key problem with this form is that each MO $|\psi_i\rangle$ is hard to interpret, as each AO's basis functions $|\phi_j\rangle$ cannot be as-

TABLE II. Numerical values of the calculations shown in Fig. 5. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full-system CCSD(T), ϵ, is given. The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for IBO localization are shown, with results for SPADE in Table III reference values using the full system given in Table I. All energies reported in hartree.

Molecule	ϵ_{Huz}	ϵ_μ	Q_{Huz}	Q_μ	$|H_{\text{emb}}^{\text{Huz}}|$	$|H_{\text{emb}}^\mu|$
N-methylmethanamine	0.526	0.526	22	22	20838	20838
acetaldehyde	0.211	0.213	28	28	54259	54011
acetonitrile	0.183	0.205	26	26	35588	34584
ethanamine	0.383	0.381	26	26	40340	40028
ethanol	0.384	0.384	24	24	29737	29737
fluoroethane	0.312	0.312	26	26	38164	38180
formamide	0.329	0.329	24	24	26769	26769

TABLE III. Numerical values of the calculations shown in Fig. 2. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full-system CCSD(T), ϵ, is given. The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for SPADE localization are shown, with results for IBO in Table II reference values using the full system given in Table I. All energies reported in hartree.

Molecule	ϵ_{Huz}	ϵ_μ	Q_{Huz}	Q_μ	$|H_{\text{emb}}^{\text{Huz}}|$	$|H_{\text{emb}}^\mu|$
N-methylmethanamine	0.342	0.342	28	28	55143	55147
acetaldehyde	0.169	0.169	30	30	71218	71338
acetonitrile	0.169	0.169	28	28	50839	50847
ethanamine	0.341	0.341	28	28	54823	54823
ethanol	0.324	0.341	26	26	41060	41068
fluoroethane	0.170	0.170	32	32	89369	89373
formamide	0.233	0.249	26	26	37008	37520

sociated with a given atom [45]. Normally MOs are highly delocalized and each $|\phi_j\rangle$ will contribute where it is needed most. Thus, we would like to expand the MOs over another minimal basis B_2 of free-atom atomic orbitals for each atom, which we write as $\{|\tilde{\phi}\rangle\} \in B_2$. This would make the wave function easy to interpret, but would be inaccurate and even incorrect, as free-atom AOs contain no polarization due to the molecular environment. The IAO method aims to combine the best properties of these.

First, a molecular SCF wave function $|\Phi\rangle$ (a single Slater determinant) is calculated that defines a set of MOs $|\psi_i\rangle$. Then a set of polarized atomic orbitals $\{|\phi^{(\text{IAO})}\rangle\} \notin B_2$ is formed that can express $|\Phi\rangle$s in the occupied MOs $|\psi_i\rangle$. To do this, projectors onto the occupied and virtual MOs are defined: $P = \sum_i |\psi_i\rangle\langle\psi_i|$ and $Q = I - P$, respectively [82]. The AO projectors onto the bases B_1 and B_2 are also defined as [45]

$$\mathbf{P}_{12} = \sum_{|\phi\rangle_i, |\phi\rangle_j \in B_1} |\phi\rangle_i \mathbf{S}_{ij}^{-1} \langle\phi|_j, \tag{A7a}$$

$$\mathbf{P}_{21} = \sum_{|\phi\rangle_k, |\phi\rangle_l \in B_2} |\phi\rangle_k \mathbf{S}_{kl}^{-1} \langle\phi|_l. \tag{A7b}$$

Here $\mathbf{S}_{ij}^{-1}$ and $\mathbf{S}_{kl}^{-1}$ are the inverse overlap matrices in the bases B_1 and B_2. The set of depolarized MOs is then given by [45,82]

$$\{|\tilde{\psi}_i\rangle\} = \{\mathbf{P}_{12}\mathbf{P}_{21}|\psi_i\rangle\}. \tag{A8}$$

These depolarized molecular orbitals are then used to define the projectors: $\tilde{\mathbf{P}} = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$ and $\tilde{\mathbf{Q}} = \mathbf{I} - \tilde{\mathbf{P}}$. The IAOs are

TABLE IV. Cartesian coordinates of atoms in H_2O for the structure with an OH bond length of 0.4 Å defined from the first H and O atoms in this table. The other structures (different OH bond lengths) were generated from this file by changing the position of the first H atom. Note the H-O-H angle remained fixed.

Atom	x	y	z
H	0.3751747	0.0000000	0.1387225
O	0.0000000	0.0000000	0.0000000
H	−0.7493682	0.0000000	0.2770822

TABLE V. Numerical values of the calculations shown in Fig. 3 for the case when the changing OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in hartree.

OH length	$E_{\text{FCI}}^{\text{glob}}$	$E_{\text{DFT}}^{\text{glob}}$	$E_{\text{FCI-in-DFT}}^{\text{Huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$|H_{\text{full}}|$	$|H_{\text{FCI-in-DFT}}^{\text{Huz}}|$	$|H_{\text{FCI-in-DFT}}^{\mu}|$	$Q_{\text{FCI}}^{\text{glob}}$	$Q_{\text{FCI-in-DFT}}^{\text{Huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	No. active MOs
0.400000	−72.981056	−73.259936	−72.988008	−72.988008	2110.000000	1079	1079	14	12	12	4
0.600000	−74.499220	−74.773419	−74.508545	−74.508545	2110.000000	1079	1079	14	12	12	4
0.798954	−74.851089	−75.122053	−74.864001	−74.864002	1086.000000	1079	1079	14	12	12	4
1.000000	−74.900658	−75.170068	−74.918225	−74.918226	2110.000000	1079	1079	14	12	12	4
1.200000	−74.867498	−75.134418	−74.890912	−74.890913	2110.000000	1079	1079	14	12	12	4
1.500000	−74.807539	−75.057383	−74.840738	−74.840739	2110.000000	1079	1079	14	12	12	4
2.000000	−74.776263	−74.962535	−74.816902	−74.816903	2110.000000	1079	1079	14	12	12	4
3.000000	−74.771826	−74.890579	−74.820209	−74.820210	2430.000000	1211	1175	14	12	12	4
4.000000	−74.771720	−74.876127	−74.820269	−74.820269	2314.000000	1819	1079	14	12	12	4
5.000000	−74.771718	−74.872360	−74.820210	−74.820210	3282.000000	1795	1795	14	12	12	4

finally given by

$$|\phi^{(\text{IAO})}\rangle = (\mathbf{P}\tilde{\mathbf{P}} + \mathbf{Q}\tilde{\mathbf{Q}})\mathbf{P}_{12}|\tilde{\phi}\rangle. \qquad (A9)$$

The IAO basis is then orthonormalized [45,82]. Overall, each IAO is obtained by a simple set of matrix operations. The utility of IAOs stems from the fact that they are directly associated with atoms and can be used to define atomic properties like partial charges and are basis-set independent, unlike the Pipek-Mezey approach.

Often we want to know about molecular bonding rather than atomic properties. Knizia showed that by combining IAOs with a localization in the spirit of Pipek-Mezey, one can obtain IBOs. As discussed in the main text, a Slater determinant $|\Phi\rangle$ is invariant to unitary rotations $|\psi_i^{\text{LMO}}\rangle = |\psi_i\rangle\mathbf{U}$; we can thus define the IBOs by maximizing [45]

$$L_{\text{IBO}} = \sum_{A}^{N_{\text{atoms}}} \sum_{i}^{N_{\text{occ}}} [n_A(i)]^4 \qquad (A10)$$

with respect to $\mathbf{U}$. Here $n_A(i)$ is the number of electrons from ψ_i^{LMO} located on the IAOs $\{|\phi^{(\text{IAO})}\rangle\}$ of atom A. Explicitly, we

write this as [45]

$$n_A(i) = 2 \sum_{|\phi^{(\text{IAO})}\rangle \in A} \langle \psi_i^{\text{LMO}}|\phi^{(\text{IAO})}\rangle\langle\phi^{(\text{IAO})}|\psi_i^{\text{LMO}}\rangle. \qquad (A11)$$

This construction minimizes the number of atoms upon which an orbital is centered [45].

In order to assign these localized molecular orbitals ($\mathbf{C}^{\text{LMO}}$) to the active and environment subsystems we calculate the percentage of the ith LMO over the active atoms as

$$p_i^{\text{act}}(\mathbf{C}) = \frac{\sum_{\nu \in \text{act}} \mathbf{C}_{i\nu}^2}{\sum_{\nu=1}^{K} \mathbf{C}_{i\nu}^2}, \qquad (A12)$$

where $\nu \in$ act are the atomic orbital indices for the atoms defined in the active region. The denominator includes all the AOs of the ith MO. This is the approach given in Eq. (10) in the work of Giovannini and Koch [83]. Any $p_i^{\text{act}} > 95\%$ we associate to the active subsystem.

Our code on GitHub uses this metric for the other localization strategies supported by PySCF. The code generates a localized $\mathbf{C}$ matrix, which can then be used in conjunction with Eq. (A12).

TABLE VI. Numerical values of the calculations shown in Fig. 3 for the case when the fixed OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in hartree.

OH length	$E_{\text{FCI}}^{\text{glob}}$	$E_{\text{DFT}}^{\text{glob}}$	$E_{\text{FCI-in-DFT}}^{\text{Huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$|H_{\text{full}}|$	$|H_{\text{FCI-in-DFT}}^{\text{Huz}}|$	$|H_{\text{FCI-in-DFT}}^{\mu}|$	$Q_{\text{FCI}}^{\text{glob}}$	$Q_{\text{FCI-in-DFT}}^{\text{Huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	No. active MOs
0.400000	−72.981056	−73.259936	−72.887821	−72.887827	2110.000000	1079	1079	14	12	12	4
0.600000	−74.499220	−74.773419	−74.473689	−74.473690	2110.000000	1079	1079	14	12	12	4
0.798954	−74.851089	−75.122053	−74.864001	−74.864002	1086.000000	1079	1079	14	12	12	4
1.000000	−74.900658	−75.170068	−74.936101	−74.936101	2110.000000	1079	1079	14	12	12	4
1.200000	−74.867498	−75.134418	−74.914336	−74.914336	2110.000000	1079	1079	14	12	12	4
1.500000	−74.807539	−75.057383	−74.847327	−74.847327	2110.000000	1079	1079	14	12	12	4
2.000000	−74.776263	−74.962535	−74.755507	−74.755507	2110.000000	1079	1079	14	12	12	4
3.000000	−74.771826	−74.890579	−74.683921	−74.683921	2250.000000	1095	1119	14	12	12	4
4.000000	−74.771720	−74.874152	−74.667467	−74.667467	2170.000000	1159	1159	14	12	12	4
5.000000	−74.771718	−74.871551	−74.667841	−74.667841	3062.000000	1795	1803	14	12	12	4

TABLE VII. Numerical values of the calculations when the changing OH bond is set as the active region and IBO localization (95% threshold) has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in hartree.

OH length	$E_{\text{FCI}}^{\text{glob}}$	$E_{\text{DFT}}^{\text{glob}}$	$E_{\text{FCI-in-DFT}}^{\text{Huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$|H_{\text{full}}|$	$|H_{\text{FCI-in-DFT}}^{\text{Huz}}|$	$|H_{\text{FCI-in-DFT}}^{\mu}|$	$Q_{\text{FCI}}^{\text{glob}}$	$Q_{\text{FCI-in-DFT}}^{\text{Huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	No. active MOs
0.400000	−72.981056	−73.259936	−72.935486	−72.923051	2110.000000	492	492	14	10	10	3
0.600000	−74.499220	−74.773419	−74.484545	−74.484550	2110.000000	492	492	14	10	10	3
0.798954	−74.851089	−75.122053	−74.852613	−74.852615	1086.000000	1079	1079	14	12	12	4
1.000000	−74.900658	−75.170068	−74.899540	−74.899542	2110.000000	1079	1079	14	12	12	4
1.200000	−74.867498	−75.134418	−74.868096	−74.868099	2110.000000	1079	1079	14	12	12	4
1.500000	−74.807539	−75.057383	−74.819774	−74.819777	2110.000000	1079	1079	14	12	12	4
2.000000	−74.776263	−74.962535	−74.798954	−74.798956	2110.000000	1079	1079	14	12	12	4
3.000000	−74.771826	−74.890579	−74.804435	−74.804443	2430.000000	876	876	14	10	10	3
4.000000	−74.771720	−74.876127	−74.805034	−74.805037	2314.000000	1815	1815	14	12	12	4
5.000000	−74.771718	−74.872360	−74.804966	−74.804968	3282.000000	1759	1759	14	12	12	4

APPENDIX B: EMBEDDED SELF-CONSISTENT FIELD METHODOLOGY

In this section we summarize how the embedded self-consistent field calculations are performed for the μ-shift and Huzinaga methods. We consider restricted Hartree-Fock (RHF) calculations; extensions to the unrestricted case follow straightforwardly.

1. μ-shift RHF

To perform the μ-shifted RHF calculation, only the core Hamiltonian is modified [Eq. (16b) in the main text], by adding $\mathbf{V}_{\text{emb}} + \mathbf{P}_{\text{proj}}^{\text{env}}$ to it. The $\mathbf{g}(\gamma^{\text{act}} + \gamma^{\text{env}})$ term is obtained from the global DFT calculation. The $\mathbf{g}(\gamma^{\text{act}})$ potential energy matrix is calculated in the same way, except the density matrix is set to be that of the active system only. The μ-shift projector can then be defined using Eq. (17) in the main text.

The standard RHF algorithm can then be run, where the only difference is that the standard core Hamiltonian has been modified to be $\mathbf{h}_{\text{emb}}$.

2. Huzinaga RHF

The Huzinaga RHF calculation is slightly more involved than the μ-shifted method. At each self-consistent field loop, when the new Fock matrix is defined, the Huzinaga projection operator $\mathbf{P}_{\text{Huz}}^{\text{env}}$ must be built according to Eq. (20) in the main text. We reiterate that this step uses the current Fock matrix in the self-consistent field loop. The embedded Fock matrix is then constructed according to Eq. (15) of the main text, where the potential energy matrix $\mathbf{V}_{\text{emb}}^{\text{proj}}$ [Eq. (16a) in the main text] is defined in the same manor as the μ-shift RHF method. Given this embedded Fock matrix, $\mathbf{F}_{\text{emb}}^{\text{act}}\mathbf{C} = \mathbf{SC}\epsilon$ can be solved via standard SCF approaches. This generates a set of new MO coefficients $\mathbf{C}$ that are used to construct the new Fock matrix. This process is repeated until the energy converges as usual in an SCF calculation.

APPENDIX C: ACTIVE ATOM SELECTION

As the number of active atoms is configurable with this method, we demonstrate the effect of altering this parameter using cyclopentane. Figure 4 shows the change in calculated

TABLE VIII. Numerical values of the calculations when the fixed OH bond is set as the active region and IBO localization (95% threshold) has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits, Q, and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in hartree.

OH length	$E_{\text{FCI}}^{\text{glob}}$	$E_{\text{DFT}}^{\text{glob}}$	$E_{\text{FCI-in-DFT}}^{\text{Huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$|H_{\text{full}}|$	$|H_{\text{FCI-in-DFT}}^{\text{Huz}}|$	$|H_{\text{FCI-in-DFT}}^{\mu}|$	$Q_{\text{FCI}}^{\text{glob}}$	$Q_{\text{FCI-in-DFT}}^{\text{Huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	No. active MOs
0.400000	−72.981056	−73.259936	−73.002214	−73.002219	2110.000000	1079	1079	14	12	12	4
0.600000	−74.499220	−74.773419	−74.519672	−74.519674	2110.000000	1079	1079	14	12	12	4
0.798954	−74.851089	−75.122053	−74.852613	−74.852615	1086.000000	1079	1079	14	12	12	4
1.000000	−74.900658	−75.170068	−74.927934	−74.927934	2110.000000	1079	1079	14	12	12	4
1.200000	−74.867498	−75.134418	−74.911230	−74.911230	2110.000000	1079	1079	14	12	12	4
1.500000	−74.807539	−75.057383	−74.847005	−74.847005	2110.000000	1079	1079	14	12	12	4
2.000000	−74.776263	−74.962535	−74.755516	−74.755516	2110.000000	1079	1079	14	12	12	4
3.000000	−74.771826	−74.890579	−74.683914	−74.683914	2250.000000	1095	1119	14	12	12	4
4.000000	−74.771720	−74.874152	−74.667459	−74.667459	2170.000000	1159	1159	14	12	12	4
5.000000	−74.771718	−74.871551	−74.664686	−74.664686	3062.000000	1795	1803	14	12	12	4

ground-state energy, qubit count, and number of terms in the Jordan-Wigner encoded qubit Hamiltonian. Note that the parameterless SPADE method may return a smaller active region for an increased number of active atoms. As the partition is made at the point of greatest change in singular value it is not necessarily true that resource requirements increase monotonically as they do for population-based methods such as IBO.

APPENDIX D: MOLECULAR GROUND-STATE ENERGY

Our method provides flexibility to select a localization method. In addition to the results displayed in Fig. 2, which were calculated using the SPADE projection method, we present results for the same molecules using the intrinsic bonding orbitals localization method in Fig. 5. Numerical values for these results are given in Table I for reference values, and Tables II and III for our calculated results. Note that the qubit numbers in Tables II and III differ due to differing numbers of active MOs. The 95% threshold to select IBO LMOs can be lowered to include more in a given embedded calculation, increasing the number of qubits in order to match those of SPADE. An example of this is given in Appendix E.

APPENDIX E: STRONG CORRELATION

We provide the numerical details of our strongly correlated H_2O study in this section, where SPADE localization

has been used. These results form Fig. 3 in the main text. We also include results for IBO localization. The raw data for all these results are provided in the Supplemental Material [84].

For the H_2O projection-based embedding calculations, at different molecular geometries, we considered two different active regions. One had the atoms in the fixed OH bond set active and the other had the atoms in the changing OH bond set active. The structure for H_2O with an OH bond length of 0.4 Å is given in Table IV. The other geometries can be generated from this structure. Tables V and VI summarize the numerical results for the different active systems where SPADE localization has been used. Tables VII and VIII give the numerical results for the different active systems when IBO localization was used and Fig. 6 provides a graphical summary of these results.

For the FCI-in-DFT results given in Fig. 6, we note that the error in the embedded calculation is actually higher than the global DFT calculation at an OH bond length of 4.0 Å. There is also a significant difference in the number of qubits and Hamiltonian terms at OH bond lengths of 0.4, 0.6, and 4.0 Å. The variation, compared with the SPADE results (Fig. 3 in the main text), is mainly due to the localization method giving different numbers of active MOs. We repeated the IBO calculation using a different active region threshold (minimum setting of 90%) and obtained similar results, as the number of active MOs then matched that of the SPADE calculation. Figure 7 summarizes this result.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] B. A. Cordier, N. P. Sawaya, G. G. Guerreschi, and S. K. McWeeney, Biology and medicine in the landscape of quantum advantages, J. R. Soc. Interface **19**, 20220541 (2022)

[3] H.-P. Cheng, E. Deumens, J. K. Freericks, C. Li, and B. A. Sanders, Application of quantum computing to biochemical systems: A look to the future, Front. Chem. **8**, 587143 (2020).

[4] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[5] Q. Sun, G. Kin, and L. Chan, Quantum embedding theories, Acc. Chem. Res. **29**, 43 (2016).

[6] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. **120**, 12685 (2020).

[7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed. (Dover, New York, 2012).

[8] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys. **79**, 291 (2007).

[9] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Sci. Technol. **4**, 014008 (2018).

[10] G. D. Purvis III and R. J. Bartlett, A full coupled-cluster singles and doubles model: The inclusion of disconnected triples, J. Chem. Phys. **76**, 1910 (1982).

[11] J. M. Herbert, Fantasy versus reality in fragment-based quantum chemistry, J. Chem. Phys. **151**, 170901 (2019).

[12] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).

[13] A. Yu. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[14] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Scalable quantum simulation of molecular energies, Phys. Rev. X **6**, 031007 (2016).

[15] H. Mohammadbagherpoor, Y.-H. Oh, A. Singh, X. Yu, and A. J. Rindos, Experimental challenges of implementing quantum phase estimation algorithms on IBM quantum computer, arXiv:1903.07605.

[16] S. Lee, J. Lee, H. Zhai *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, Nat. Commun. **14**, 1952 (2023).

[17] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[18] Google AI Quantum and Collaborators, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo,

M. Broughton, B. B. Buckley *et al.*, Hartree-Fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[19] L. Zhao, J. Goings, K. Wright, J. Nguyen, J. Kim, S. Johri, K. Shin, W. Kyoung, J. I. Fuks, J.-K. K. Rhee *et al.*, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, npj Quantum Inf. **9**, 60 (2023).

[20] M. Rossmannek, P. K. Barkoutsos, P. J. Ollitrault, and I. Tavernelli, Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems, J. Chem. Phys. **154**, 114105 (2021).

[21] H. Ma, M. Govoni, and G. Galli, Quantum simulations of materials on near-term quantum computers, npj Comput. Mater. **6**, 85 (2020).

[22] G. Knizia and G. K.-L. Chan, Density matrix embedding: A simple alternative to dynamical mean-field theory, Phys. Rev. Lett. **109**, 186404 (2012).

[23] N. C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory, arXiv:1610.06910.

[24] T. Yamazaki, S. Matsuura, A. Narimani, A. Saidmuradov, and A. Zaribafiyan, Towards the practical application of near-term quantum computers in quantum chemistry simulations: A problem decomposition approach, arXiv:1806.01305.

[25] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid quantum-classical approach to correlated materials, Phys. Rev. X **6**, 031045 (2016).

[26] J. M. Kreula, L. García-Álvarez, L. Lamata, S. R. Clark, E. Solano, and D. Jaksch, Few-qubit quantum-classical simulation of strongly correlated lattice fermions, EPJ Quantum Technol. **3**, 11 (2016).

[27] T. Steckmann, T. Keen, A. F. Kemper, E. F. Dumitrescu, and Y. Wang, Simulating the Mott transition on a noisy digital quantum computer via Cartan-based fast-forwarding circuits, arXiv:2112.05688.

[28] N. C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory, arXiv:1610.06910.

[29] H. Ma, M. Govoni, and G. Galli, Quantum simulations of materials on near-term quantum computers, npj Comput. Mater. **6**, 1 (2020).

[30] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, Embedding methods for quantum chemistry: Applications from materials to life sciences, J. Am. Chem. Soc. **142**, 3281 (2020).

[31] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller, A simple, exact density-functional-theory embedding scheme, J. Chem. Theory Comput. **8**, 2564 (2012).

[32] S. Lehtola and H. Jónsson, Unitary optimization of localized molecular orbitals, J. Chem. Theory Comput. **9**, 5365 (2013).

[33] D. Claudino and N. J. Mayhall, Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories, J. Chem. Theory Comput. **15**, 1053 (2019).

[34] S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller, Projection-based wavefunction-in-DFT embedding, Acc. Chem. Res. **52**, 1359 (2019).

[35] O. Roncero, M. de Lara-Castells, P. Villarreal, F. Flores, J. Ortega, M. Paniagua, and A. Aguado, An inversion technique for the calculation of embedding potentials, J. Chem. Phys. **129**, 184104 (2008).

[36] T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, Accurate basis set truncation for wavefunction embedding, J. Chem. Phys. **139**, 024103 (2013).

[37] B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, Exact density functional and wave function embedding schemes based on orbital localization, J. Chem. Phys. **145**, 064107 (2016).

[38] S. Huzinaga and A. A. Cantu, Theory of separability of many-electron systems, J. Chem. Phys. **55**, 5543 (1971).

[39] E. Francisco, A. Martín Pendás, and W. Adams, Generalized Huzinaga building-block equations for nonorthogonal electronic groups: Relation to the Adams-Gilbert theory, J. Chem. Phys. **97**, 6504 (1992).

[40] T. Shimazaki, K. Kitaura, D. G. Fedorov, and T. Nakajima, Group molecular orbital approach to solve the Huzinaga subsystem self-consistent-field equations, J. Chem. Phys. **146**, 084109 (2017).

[41] D. V. Chulhai and J. D. Goodpaster, Projection-based correlated wave function in density functional theory embedding for periodic systems, J. Chem. Theory Comput. **14**, 1928 (2018).

[42] D. V. Chulhai and J. D. Goodpaster, Improved accuracy and efficiency in quantum embedding through absolute localization, J. Chem. Theory Comput. **13**, 1503 (2017).

[43] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, Accurate and systematically improvable density functional theory embedding for correlated wavefunctions, J. Chem. Phys. **140**, 18A507 (2014).

[44] D. S. Graham, X. Wen, D. V. Chulhai, and J. D. Goodpaster, Huzinaga projection embedding for efficient and accurate energies of systems with localized spin-densities, J. Chem. Phys. **156**, 054112 (2022).

[45] G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts, J. Chem. Theory Comput. **9**, 4834 (2013).

[46] A. E. Reed, R. B. Weinstock, and F. Weinhold, Natural population analysis, J. Chem. Phys. **83**, 735 (1985).

[47] J. Pipek and P. G. Mezey, A fast intrinsic localization procedure applicable for *ab initio* and semi-empirical linear combination of atomic orbital wave functions, J. Chem. Phys. **90**, 4916 (1989).

[48] R. S. Mulliken, Electronic population analysis on LCAO-MO molecular wave functions. I, J. Chem. Phys. **23**, 1833 (1955).

[49] J. Foster and S. Boys, Canonical configurational interaction procedure, Rev. Mod. Phys. **32**, 300 (1960).

[50] C. Edmiston and K. Ruedenberg, Localized atomic and molecular orbitals, Rev. Mod. Phys. **35**, 457 (1963).

[51] I.-M. Høyvik, B. Jansik, and P. Jørgensen, Orbital localization using fourth central moment minimization, J. Chem. Phys. **137**, 224114 (2012).

[52] M. Williams de la Bastida and A. Ralli, Nbed, https://github.com/UCL-CCS/Nbed.

[53] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma *et al.*, Pyscf: The PYTHON-based simulations of chemistry framework, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).

[54] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle *et al.*, OpenFermion: The electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[55] P. Jordan and E. Wigner, Uber das Paulische Aquivalenzverbot, Z. Phys. **47**, 631 (1928).

[56] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, Pubchem in 2021: New data content and improved web interfaces, Nucl. Acids Res. **49**, D1388 (2021).

[57] H. Bergwerf, Molview, https://molview.org/.

[58] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush *et al.*, Quantum chemistry calculations on a trapped-ion quantum simulator, Phys. Rev. X **8**, 031022 (2018).

[59] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[60] A. D. Bochevarov and R. A. Friesner, The densities produced by the density functional theory: Comparison to full configuration interaction, J. Chem. Phys. **128**, 034102 (2008).

[61] X. Yuan, L. Visscher, and A. S. P. Gomes, Assessing MP2 frozen natural orbitals in relativistic correlated electronic structure calculations, J. Chem. Phys. **156**, 224108 (2022).

[62] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents? J. Chem. Phys. **152**, 124107 (2020).

[63] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, Doubling the size of quantum simulators by entanglement forging, PRX Quantum **3**, 010309 (2022).

[64] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Insights into current limitations of density functional theory, Science **321**, 792 (2008).

[65] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Challenges for density functional theory, Chem. Rev. **112**, 289 (2012).

[66] T. Van Voorhis and M. Head-Gordon, Benchmark variational coupled cluster doubles results, J. Chem. Phys. **113**, 8873 (2000).

[67] S. M. Sharada, D. Stück, E. J. Sundstrom, A. T. Bell, and M. Head-Gordon, Wavefunction stability analysis without analytical electronic Hessians: Application to orbital-optimised second-order Møller-Plesset theory and VV10-containing density functionals, Mol. Phys. **113**, 1802 (2015).

[68] D. Hait, N. M. Tubman, D. S. Levine, K. B. Whaley, and M. Head-Gordon, What levels of coupled cluster theory are appropriate for transition metal systems? A study using near-exact quantum chemical values for $3d$ transition metal binary compounds, J. Chem. Theory Comput. **15**, 5370 (2019).

[69] J. Shee, M. Loipersberger, D. Hait, J. Lee, and M. Head-Gordon, Revealing the nature of electron correlation in transition metal complexes with symmetry breaking and chemical intuition, J. Chem. Phys. **154**, 194109 (2021).

[70] T. J. Lee and P. R. Taylor, A diagnostic for determining the quality of single-reference electron correlation methods, Int. J. Quantum Chem. **36**, 199 (1989).

[71] C. L. Janssen and I. M. Nielsen, New diagnostics for coupled-cluster and Møller-Plesset perturbation theory, Chem. Phys. Lett. **290**, 423 (1998).

[72] U. R. Fogueri, S. Kozuch, A. Karton, and J. M. Martin, A simple DFT-based diagnostic for nondynamical correlation, Theor. Chem. Acc. **132**, 1291 (2013).

[73] M. Fuchs, Y.-M. Niquet, X. Gonze, and K. Burke, Describing static correlation in bond dissociation by Kohn-Sham density functional theory, J. Chem. Phys. **122**, 094116 (2005).

[74] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for Diels-Alder reactions and Pt(P(*t*-Bu)$_3$)$_2$+ H$_2$ oxidative addition, J. Phys. Chem. **100**, 19357 (1996).

[75] W. M. Kirby and P. J. Love, Classical simulation of noncontextual Pauli Hamiltonians, Phys. Rev. A **102**, 032418 (2020).

[76] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, arXiv:1701.08213.

[77] M. D. Radin and P. Johnson, Classically-boosted variational quantum eigensolver, arXiv:2106.04755.

[78] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, Phys. Rev. Lett. **122**, 140504 (2019).

[79] G. Wang, D. E. Koh, P. D. Johnson, and Y. Cao, Minimizing estimation runtime on noisy quantum computers, PRX Quantum **2**, 010346 (2021).

[80] J. P. T. Stenger, D. Gunlycke, and C. S. Hellberg, Expanding variational quantum eigensolvers to larger systems by dividing the calculations between classical and quantum hardware, Phys. Rev. A **105**, 022438 (2022).

[81] D. Claudino and N. J. Mayhall, Simple and efficient truncation of virtual spaces in embedded wave functions via concentric localization, J. Chem. Theory Comput. **15**, 6085 (2019).

[82] S. Barison, D. E. Galli, and M. Motta, Quantum simulations of molecular systems with intrinsic atomic orbitals, Phys. Rev. A **106**, 022404 (2022).

[83] T. Giovannini and H. Koch, Energy-based molecular orbital localization in a specific spatial region, J. Chem. Theory Comput. **17**, 139 (2021).

[84] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.109.022418 for raw data used to generate the figures and tables in this manuscript.